

Determinantal sampling designs

V. Loonis

Insee, Division des Méthodes et des Référentiels Géographiques, Paris, France.

X. Mary

Université Paris-Ouest Nanterre-La Défense, Laboratoire MODAL'X, Nanterre, France.

Summary. In this article, recent results about point processes are imported in sampling theory. Precisely, we define and study a new class of sampling designs: determinantal sampling designs. The law of such designs is known, and there exists a simple selection algorithm. We compute exactly the variance of linear estimators constructed upon those designs by using the first and second order inclusion probabilities. Moreover, we obtain asymptotic and finite distance theorems. We also adress the search of optimal determinantal sampling designs.

Keywords: sampling designs, Horvitz-Thompson estimator, determinantal point processes, balanced sampling designs.

The goal of sampling theory is to acquire knowledge of a parameter of interest θ using only partial information. The parameter θ is a function of $\{y_k, k \in U\}$, the set of values of a variable of interest y on a finite population U of size N , but the variable y is only observed on a sample $s \subseteq U$. One therefore replaces θ by $\hat{\theta}$, function of the values $\{y_k, k \in s\}$ on the sample s . $\hat{\theta}$ is called an estimator of θ , and its statistical properties can be studied as soon as the sample is randomly chosen. We then denote by \mathbb{S} a random variable with values in all the possible samples 2^U , the set of parts of U . The probability that a particular element k belongs to the randomly chosen sample, $pr(k \in \mathbb{S})$ is called the first order inclusion probability and we note $\pi_k = pr(k \in \mathbb{S})$. The second order inclusion probability $pr((k, l) \in \mathbb{S})$ is the probability that the two elements k and l both belong to the selected sample. We note $\pi_{kl} = pr((k, l) \in \mathbb{S})$ for $k \neq l$ and $\pi_{kk} = \pi_k = pr(k \in \mathbb{S})$ (for sampling without replacement, $pr((k, k) \in \mathbb{S}) = 0$). The law of \mathbb{S} is called the sampling design. It can be “universal” (as simple random sampling), but is usually chosen regarding auxiliary variables x .

Typical parameters to estimate are the total $t_y = \sum_{k \in U} y_k$, sum of the values of the variable of interest y over the whole population, or the mean value t_y/N . An estimator of t_y is called linear and homogeneous if there exist weights $w_k(\mathbb{S}), k \in U$ (that may depend on the sample) such that the estimator writes

$$\hat{t}_{yw} = \sum_{k \in \mathbb{S}} w_k(\mathbb{S}) y_k.$$

Let y be the vector $(y_1, \dots, y_N)^T$. When the weights do not depend on the sample, then the Mean Square Error (MSE) decomposes as:

$$\text{MSE}(\hat{t}_{yw}) = \overbrace{\sum_{k \in U} \sum_{l \in U} w_k w_l y_k y_l \Delta_{kl}}^{\text{Variance}} + \left[\overbrace{\sum_{k \in U} (w_k \pi_k - 1) y_k}^{\text{Bias}} \right]^2 \quad (1)$$

where

$$\Delta_{kl} = \begin{cases} \pi_{kl} - \pi_k \pi_l & (k \neq l), \\ \pi_k(1 - \pi_k) & (k = l). \end{cases}$$

The Horvitz-Thompson estimator (or Narain-Horvitz-Thompson estimator Narain (1951), Horvitz (1952)) \hat{t}_{yHT} is the best known of all the homogeneous linear estimators of a total. Its weights do not depend on the sample and satisfy $w_k = \pi_k^{-1}$ for positive π_k and w_k arbitrarily chosen otherwise. Statistical properties such as expectancy, variance or consistency \hat{t}_{yw} highly depend on the first or second order inclusion probabilities of the sampling procedure. For instance, the Horvitz-Thompson estimator is the only unbiased linear estimator, whose weights do not depend on \mathbb{S} , if and only if (iff) all first order probabilities are positive. If moreover the second order probabilities are positive, then the sampling variance may itself be unbiasedly estimated by means of the Horvitz-Thompson variance estimator, or in case of fixed-size sampling designs, by the Sen-Yates-Grundy variance estimator (Hedayat (1991)). However, the second order inclusion probabilities are hardly known (or at prohibitive costs) in practice. The existence of a large, flexible family of sampling designs with known first and second order inclusion probabilities is then of particular interest.

In this article, we introduce a new parametric family of sampling designs, determinantal sampling designs, whose inclusion probabilities are known for any order. Section 1 gives their definition, probabilistic properties and also provides a selection algorithm. Section 2 focuses on the existence and construction of determinantal sampling designs with fixed first order inclusion probabilities. Section 3 studies the statistical properties of linear estimators of a total, for determinantal sampling designs. In this section, we first give algebraic and geometric formulas for the MSE which enable us to provide necessary and sufficient conditions for a perfect balanced determinantal sampling design. In a second time we give asymptotic theorems and concentration inequalities. Section 4 addresses the question of optimality. Finally, Section 5 presents the results of a study of a concrete case.

1. Determinantal sampling designs

In this section, we define our object of study: determinantal sampling designs. We focus notably on two important features: inclusion probabilities and sample size. Other probabilistic properties are also interpreted in the sampling framework. Finally we provide a sampling algorithm.

1.1. Definitions

Recall the next two definitions. The first one comes from sampling theory (Sarndal (1992)), whereas the second one comes from the theory of probability (Borodin (2011)).

DEFINITION 1.1. *An unordered sampling design without replacement \mathcal{P} on a finite set U is a probability on 2^U , set of parts of U . We say that a random variable \mathbb{S} is drawn from \mathcal{P} if its image measure is \mathcal{P} , that is $pr(\mathbb{S} = s) = \mathcal{P}(s)$ for all $s \in 2^U$.*

DEFINITION 1.2. *A simple point process \mathcal{P} on U is a probability on 2^U .*

This is the starting point of this article: an unordered sampling design without replacement is exactly a simple point process on a finite set.

Among simple point processes, the general structure and properties of determinantal point processes have attracted a lot of attention recently (Borodin (2011), Hough (2006), Hough (2009), Lyons (2003), Soshnikov (2000)). This is in part due to the ubiquity of determinantal point processes in probability theory. They appear for instance in the study of random structures such as uniform spanning trees, zeros of random polynomials and random matrices. In the case of a finite set U , determinantal point processes are defined through associated matrices called kernels (such a kernel is however not unique). Many probabilistic properties of these processes therefore depend on algebraic properties of their kernels, but most of the results concern Hermitian matrices only. For this reason, and though there exist many interesting examples of determinantal point processes associated to non-Hermitian matrices, we restrict our attention to the Hermitian case. From the definition of determinantal point processes we propose the following definition of determinantal sampling designs:

DEFINITION 1.3 (DETERMINANTAL SAMPLING DESIGN). *A sampling design \mathcal{P} on a finite set U is a determinantal sampling design if there exists an Hermitian matrix K indexed by U , called kernel, such that for all $s \in 2^U \setminus \emptyset$ (non-empty parts of 2^U):*

$$\sum_{s' \supseteq s} \mathcal{P}(s') = \det(K|_s)$$

where $K|_s$ denotes the submatrix of K whose rows and columns are indexed by s . This sampling design is denoted by \mathcal{P}_K .

A random variable \mathbb{S} with values in 2^U and law \mathcal{P}_K is called a determinantal random sample (with kernel K). It satisfies for all $s \in 2^U \setminus \emptyset$:

$$pr(s \subseteq \mathbb{S}) = \det(K|_s).$$

In the following we will always identify population U of size N with $\{1, \dots, N\}$. It is clear from the definition that determinantal sampling designs are unordered and without replacement. Macchi (1975) and Soshnikov (2000) proved that \mathcal{P}_K exists iff K is a contracting matrix, that is a Hermitian matrix whose eigenvalues are in $[0, 1]$. It follows from this fundamental result that determinantal sampling designs form a parametric family of sampling

designs, parametrized by contracting matrices. Because these matrices are a pivotal notion regarding determinantal sampling designs, and are less known than positive semidefinite matrices, we give some of their various characterizations below. The last characterization uses finite dimensional frames, for which we refer to Christensen (2003). These results of linear algebra will be thoroughly used in the sequel.

PROPOSITION 1.1. *Let K be a matrix of size N over \mathbb{C} . The following statements are equivalent:*

- (a) *K is a contracting matrix of rank n , that is a Hermitian matrix of rank n with eigenvalues in $[0, 1]$.*
- (b) *K is a Hermitian matrix of rank n that satisfies $0 \leq K \leq I_N$ for the Loewner order, that is K and $I_N - K$ are positive semidefinite.*
- (c) *There exists an orthogonal projection matrix P ($P\bar{P}^T = P$) and a diagonal matrix D with exactly n non-zero coefficients in $]0, 1]$ such that $K = PDP^{-1}$ (spectral decomposition).*
- (d) *There exists an orthonormal family $\{v_1, \dots, v_n\}$ of \mathbb{C}^N and a family $\{\lambda_1, \dots, \lambda_n\}$ of numbers in $]0, 1]$ such that $K = \sum_{i=1}^n \lambda_i v_i \bar{v}_i^T$ (rank one decomposition).*
- (e) *There exists a family $\{f_1, \dots, f_N\}$ of vectors of \mathbb{C}^n such that K is the Gramian matrix of (f_1, \dots, f_N) ($K_{kl} = \bar{f}_k^T f_l$) and $\{f_1, \dots, f_N\}$ is a frame of \mathbb{C}^n with upper bound $B \leq 1$ (frame decomposition).*

In matrix form, statement d) is equivalent with the existence of a matrix V of size $N \times n$ and a $n \times n$ diagonal matrix D with diagonal coefficients in $]0, 1]$ such that $\bar{V}^T V = I_n$ and $K = VD\bar{V}^T$.

Statement e) is equivalent with the existence of a matrix F of size $n \times N$ such that $K = \bar{F}^T F$ and $0 < F\bar{F}^T \leq I_n$ for the Loewner order.

Let us give now a few typical examples.

EXAMPLE 1.1 (PROJECTION). *Let P be a Hermitian projection matrix. Then $P = \bar{P}^T$ and $P^2 = P$, hence P is an orthogonal projection matrix. Therefore, we will make no distinction between projections and orthogonal projections. As the eigenvalues of P are 0 or 1, then P is a contracting matrix. We can thus associate to P a determinantal sampling design \mathcal{P}_P . We will see that \mathcal{P}_P enjoys interesting statistical properties. Such determinantal point processes are sometimes called determinantal projection processes (Hough (2006)) or elementary determinantal point processes (Kuleska (2011)) in the litterature.*

Among these sampling designs, we may single out 3 particular cases. Let J_N be the square matrix of size N with all terms equal to 1.

- (a) *The sampling design associated to $\frac{1}{N}J_N$ is the simple random sampling (SRS) of size 1;*
- (b) *The sampling design associated to $I_N - \frac{1}{N}J_N$ is the SRS of size $N - 1$. We will see later that the other SRS are non-determinantal.*
- (c) *If P is a diagonal projection matrix, the associated sampling designs is a degenerated one (non-random).*

EXAMPLE 1.2 (PROJECTION (2)). Let $n \leq N$ be integers. We will encounter in Section 2.1 the complex projection matrix $P^{1,N,n}$ defined by

$$\begin{cases} P_{kk}^{1,N,n} = \frac{n}{N}, \\ P_{kl}^{1,N,n} = \frac{1}{N} \frac{\sin\left(\frac{n(k-l)\pi}{N}\right)}{\sin\left(\frac{(k-l)\pi}{N}\right)} e^{i\left(\frac{(n-1)(k-l)\pi}{N}\right)} \quad (k \neq l). \end{cases}$$

and see that $\mathcal{P}_{P^{1,N,n}}$ defines an equal-weighted sampling design of fixed size n among a population of size N . For instance $P^{1,4,3}$ is the matrix

$$P^{1,4,3} = \frac{1}{4} \begin{pmatrix} 3 & -i & 1 & i \\ i & 3 & -i & 1 \\ 1 & i & 3 & -i \\ -i & 1 & i & 3 \end{pmatrix}.$$

EXAMPLE 1.3 (PROJECTION (3)). Let A be any (N, p) matrix of rank n . Then matrix $P_A = A(\bar{A}^T A)^\dagger \bar{A}^T$ is an orthogonal projection matrix of rank n . (where B^\dagger is the Moore-Penrose inverse of B (Ben-Israel (2003))). This is a simple procedure to construct $N \times N$ projection matrices from arbitrary matrices.

EXAMPLE 1.4 (FRAMES AND PROJECTIONS). Let $\{f_1, \dots, f_N\}$ be a Parseval tight frame of \mathbb{C}^n (Christensen (2003)). Then the Gramian matrix K of (f_1, \dots, f_N) is a projection of rank n . Conversely, any $N \times N$ projection matrix of rank n corresponds to a Parseval tight frame of length N of \mathbb{C}^n .

EXAMPLE 1.5 (POISSON SAMPLING). Consider now diagonal matrix K with diagonal elements $K_{kk} = \Pi_k \in [0, 1]$. It is a contracting matrix and the corresponding determinantal sampling design satisfies for all $s \in 2^U$

$$pr(s \subseteq \mathbb{S}) = \prod_{k \in s} \Pi_k.$$

The inclusion-exclusion principle implies that

$$pr(\mathbb{S} = s) = \prod_{k \in s} \Pi_k \prod_{k \notin s} (1 - \Pi_k).$$

This is precisely the equation of the Poisson sampling design (with first order inclusion probabilities $\pi_k = \Pi_k$), which therefore belongs to the family of determinantal sampling designs.

EXAMPLE 1.6 (L-SAMPLING DESIGN). Let L be a positive semidefinite matrix. Then matrix $K = L(I_N + L)^{-1}$ is a contracting matrix. Macchi (1975) proved that the associated determinantal process \mathcal{P}_K satisfies for all non-empty sets $s \in 2^U \setminus \emptyset$:

$$pr(\mathbb{S} = s) = \frac{\det(L|_s)}{\det(I_N + L)}.$$

It is called a L -sampling design.

1.2. Inclusion probabilities

The following formulas for the inclusion probabilities of order 1 and 2 follow from Definition 1.3. For all $k, l \in U$,

$$\pi_k = K_{kk}, \quad (2)$$

$$\pi_{kl} = K_{kk}K_{ll} - |K_{kl}|^2 \quad (k \neq l), \quad (3)$$

$$\Delta_{kl} = \begin{cases} -|K_{kl}|^2 & (k \neq l), \\ K_{kk}(1 - K_{kk}) & (k = l). \end{cases} \quad (4)$$

or in matrix formulation

$$\Delta = \overline{(I_N - K)} * K = (I_N - K) * \overline{K} \quad (5)$$

where $*$ is the Schur-Hadamard (entrywise) matrix product.

EXAMPLE 1.7 (LAPLACE KERNEL). *Set $0 < \alpha < 1$ and let $x \in \mathbb{R}^Q$ be an auxiliary variable. For β large enough, there exists a determinantal sampling design with first and second order inclusion probabilities*

$$\begin{cases} \pi_k &= \alpha, \\ \pi_{kl} &= \alpha^2(1 - \exp^{-2\beta(|x_k - x_l|)}) \quad (k \neq l). \end{cases}$$

Indeed, the Laplace kernel $f^{\alpha, \beta}(x, y) = \alpha \exp^{-\beta|x-y|}$ is positive semidefinite on \mathbb{R}^Q with $\alpha > 0$ et $\beta > 0$. By submersion, the matrix $L^{\alpha, \beta}$ defined by $L_{kl}^{\alpha, \beta} = f^{\alpha, \beta}(x_k, x_l)$ is also positive-definite. For β large enough, its eigenvalues are less than 1. The quantity αN is the average size of the random sample. If $x_k = x_l$, then k and l will never be sampled simultaneously.

EXAMPLE 1.8 (PROJECTION (2)). *Determinantal sampling design \mathcal{P}_P of Example 1.2 satisfies*

$$\begin{cases} \pi_k &= \frac{n}{N}, \\ \pi_{kl} &= \frac{n^2}{N^2} - \frac{1}{N^2} \frac{\sin^2\left(\frac{n(k-l)\pi}{N}\right)}{\sin^2\left(\frac{(k-l)\pi}{N}\right)} \quad (k \neq l). \end{cases}$$

EXAMPLE 1.9 (FRAME). *Let $\{f_1, \dots, f_N\}$ be a frame of \mathbb{C}^n with upper bound less than 1. By Proposition 1.1, its Gramian matrix K is a contracting matrix. Let \mathcal{P}_K be the associated Determinantal sampling design. Then*

$$\begin{cases} \pi_k &= \|f_k\|^2 \text{ square norm of vector } f_k, \\ \pi_{kl} &= \|f_k\|^2 \|f_l\|^2 - |\langle f_k, f_l \rangle|^2 \\ &\quad \text{area of the parallelogram determined by } f_k \text{ and } f_l \quad (k \neq l). \end{cases}$$

PROPOSITION 1.2. *From (4) a determinantal sampling design satisfies the so-called Sen-Yates-Grundy conditions:*

$$\pi_{kl} \leq \pi_k \pi_l \quad (k \neq l). \quad (6)$$

More generally, a determinantal sampling design has negative associations (Lyons (2003)). It satisfies in particular for disjoint subsets A and B

$$pr(A \cup B \subseteq \mathbb{S}) \leq pr(A \subseteq \mathbb{S})pr(B \subseteq \mathbb{S})$$

It was shown recently that determinantal point processes actually enjoy the strong Rayleigh property (Borcea (2009), Pemantle and Peres (2014)), a technical property stronger than negative association. This property can be defined in terms of the localisation of the zeros of the generating function of the process. These two properties (negative association, strong Rayleigh property) proved very useful for the study of statistics of determinantal processes (Yuan (2003), Brandèn (2012), Pemantle and Peres (2014)). Some results will be used in Section 3.

1.3. Sample size of determinantal designs and fixed size sampling designs

Of major importance to statisticians is the sample size of the random sample. It is for instance very common in practice to work with fixed size samples, that is with samples whose size is non-random and given.

The sample size of a determinantal random sample follows from Theorem 7 in Hough (2006). For a set A , let $\sharp A$ denotes its cardinal and for a Hermitian matrix K , let $Sp(K) = \{\lambda_i, i \in N\}$ be the set of eigenvalues of K (with their multiplicities).

THEOREM 1 (SAMPLE SIZE). *Let \mathbb{S} be a determinantal random sample with kernel K . Then the random variable $\sharp \mathbb{S}$ has the law of a sum of N independent Bernoulli variables B_i of parameter $\lambda_i \in Sp(K)$.*

COROLLARY 1.1 (SAMPLE SIZE (2)). *Let \mathbb{S} be a determinantal random sample with kernel K . Then*

- (a) $E(\sharp \mathbb{S}) = tr(K)$.
- (b) $var(\sharp \mathbb{S}) = tr(K - K^2) = \sum_{k \in Sp(K)} \lambda_k(1 - \lambda_k) = \sum_{k, l \in U} \Delta_{kl}$.
- (c) $pr(\mathbb{S} = \emptyset) = 0$ iff $1 \in Sp(K)$.
- (d) *The determinantal sampling design \mathcal{P}_K has fixed size iff K is an orthogonal projection matrix.*

Proof Let $\pi^T = (\pi_1, \dots, \pi_N)$ be the row vector of first inclusion probabilities. Then $var(\sharp \mathbb{S}) = \sum_{k, l \in U} \Delta_{kl}$ follows from (1) and the identification $\sharp \mathbb{S} = \hat{t}_{\pi HT}$. The other results follow directly from Theorem 1 and the spectral decomposition of Hermitian matrices. \square

Recall that in case of fixed-size sampling designs we have the stronger formula: for any $k \in U$,

$$\sum_{l \in U} \Delta_{kl} = 0. \tag{7}$$

COROLLARY 1.2. *Let $L \neq 0$ be a positive semidefinite matrix. Then the L -sample design satisfies $pr(\mathbb{S} = \emptyset) > 0$. It is not of fixed size.*

Proof Let \mathcal{P} be the associated sample design. It is a determinantal sample design with kernel $K = L(I_N + L)^{-1}$. Let $X \in \mathbb{C}^N$, $KX = X$. Then $(I_N + L)KX = LX = (I_N + L)X$ and $X = 0$. It follows that 1 is not an eigenvalue of K and by the *c.* Corollary 1.1, $pr(\mathbb{S} = \emptyset) > 0$. As $L \neq 0$, $K \neq 0$ and K is not a projection. By *d.* Corollary 1.1, it is not of fixed size. \square

1.4. Additional properties

We give here some other general probabilistic results on determinantal sampling designs and their interpretation in terms of sampling theory. The probabilistic versions may for instance be found in Lyons (2003) and Hough (2006).

PROPOSITION 1.3 (COMPLEMENTARY DESIGN). *Let \mathbb{S} be a determinantal random sample with kernel K . The complementary random sample \mathbb{S}^c is a determinantal random sample with kernel $I_N - K$.*

PROPOSITION 1.4 (DOMAIN). *Let \mathcal{P}_K be a determinantal sampling design on U with kernel K , and let $A \subseteq U$ be a subpopulation (or domain). Then the restriction $\mathcal{P}_{K|_A}$ of \mathcal{P}_K to A is determinantal sampling design on A with kernel $K|_A$, the submatrix of K whose rows and columns are indexed by A .*

PROPOSITION 1.5 (STRATIFICATION). *Let $\{U_1, \dots, U_H\}$ be a partition of U into H strata. The determinantal sampling design \mathcal{P}_K with kernel K is stratified iff the matrix K admits a block diagonal decomposition relative to these strata, that is $k \in U_h, l \in U_{h'}, h \neq h'$ implies $K_{kl} = 0$.*

By using the inclusion-exclusion principle Lyons shows that the probabilities of disjunction are also given by a determinant (Theorem 5.1 Equation (5.2) Lyons (2003) for fixed-size designs and Equation (8.1) for random size designs), but the expression is given using outer products rather than determinant of matrices. We can rewrite his result as follows. Let K be a kernel on the population U . For any two subsets s, s' of U , we set $\gamma = s \dot{\cup} s'$, where $\dot{\cup}$ denotes the disjoint union of two sets (coproduct in the category of sets). We also pose $q = \sharp\gamma (= \sharp s + \sharp s')$ and define a new square matrix $B(s, s')$ of size $q \times q$ indexed by γ by: for all $k, l \in \gamma$,

$$\begin{cases} B(s, s')_{kl} = K_{kl} & \text{if } k \in s, \\ B(s, s')_{kl} = (I_N - K)_{kl} & \text{if } k \in s'. \end{cases}$$

THEOREM 2 (DISJUNCTION). *Let \mathbb{S} be a determinantal random sample on U with kernel K and s, s' be two subsets of U such that $\gamma = s \dot{\cup} s'$ is non-empty. Then*

$$pr(s \subseteq \mathbb{S}, s' \subseteq \mathbb{S}^c) = \det(B(s, s')) \quad (8)$$

A different formulation of this probability was also obtained by Soshnikov (2000) (under the name Janossy densities):

$$pr(s \subseteq \mathbb{S}, s' \subseteq \mathbb{S}^c) = \det(I_q - K|_\gamma) \det(\Gamma|_s)$$

with $\Gamma = \left((I_q - K|_\gamma)^{-1} K|_\gamma \right)$.

We deduce from (8) the entropy of determinantal sampling designs.

COROLLARY 2.1 (ENTROPY). *Let \mathcal{P}_K be a determinantal sample design with kernel K . Then*

$$\mathcal{P}_K(s) = \det(B(s, s^c))$$

and

$$H(\mathcal{P}_K) = - \sum_{s \in 2^U} \mathcal{P}_K(s) \log(\mathcal{P}_K(s)) = - \sum_{s \in 2^U} \det(B(s, s^c)) \log(\det(B(s, s^c)))$$

Finally, a determinantal random sample conditioned on containing certain elements and excluding others is still determinantal (Lyons (2003)). Moreover the kernel admits an explicit expression in the case of fixed-size determinantal samples (Proposition 6.3 and Equation (6.5) in Lyons (2003)). In practice, this is sufficient to construct a selection algorithm.

1.5. Algorithm

The selection of a determinantal random sample from a determinantal sample design can be performed using Algorithm 18 in Hough (2006), which relies on conditioning. A very close formulation and another algorithm based on the Gram-Schmidt procedure can be found in Scardicchio (2009) and Lavancier (2015), respectively.

The next algorithm describes solely in terms of matrices the selection of a random sample for a fixed-size determinantal sampling design \mathcal{P}_K of size n , on the population $U = \{1, \dots, N\}$. For any orthogonal projection matrix K , V denotes the representative matrix of an orthonormal basis of the range of K (and $K = V\bar{V}^T$). For an element $k \in U$, e_k denotes the vector of \mathbb{C}^N whose N coordinates equal 0 except $e_k(k)$ which equals 1.

ALGORITHM 1.1. *Pose $K_n = K$. Then for $i = n, \dots, 1$ iteratively*

- (a) *Select the element $k \in U$ with probability $\frac{K_i(k, k)}{i}$. Let k_i be this element;*
- (b) *Set $Z_i = \frac{\bar{V}_i^T e_{k_i} e_{k_i}^T V_i}{K_i(k_i, k_i)}$;*
- (c) *Set $K_{i-1} = V_i(I_i - Z_i)\bar{V}_i^T$.*

The sample $\{k_1, \dots, k_n\}$ is a realisation of the determinantal sampling design \mathcal{P}_K .

Next theorem describes a procedure to sample from any determinantal sampling design, by reducing it to the fixed-size case. That this procedure sample from the determinantal sampling design \mathcal{P}_K is the content of Theorem 7 in Hough (2006), which expresses any determinantal sampling design as a mixture of fixed-size sampling designs.

THEOREM 3. *Let \mathcal{P}_K be a determinantal sampling design with kernel K whose rank one decomposition is*

$$K = \sum_{i=1}^N \lambda_i v_i \bar{v}_i^T$$

where $\{v_i, i = 1, \dots, N\}$ denotes an orthonormal family of eigenvectors. Let also \mathcal{P}_{K_I} be the (conditional) fixed-size determinantal sampling design with kernel projection matrix

$$K_B = \sum_{i=1}^N B_i v_i \bar{v}_i^T$$

where $B_i, i = 1, \dots, N$, are independent Bernoulli random variables with parameter λ_i . Then

$$\mathcal{P}_K \sim \mathcal{P}_{K_B}.$$

The concrete steps of the sampling are then:

- (a) Simulate a vector b with law B , realisation of N independent Bernoulli variables with parameter λ_i .
- (b) Construct the projection matrix $K_b = \sum_{i=1}^N b_i v_i \bar{v}_i^T$.
- (c) Sample from \mathcal{P}_{K_b} by Algorithm 1.1

Theorem 3 guarantees that this sample is a realisation of \mathcal{P}_K .

2. Determinantal sampling designs with fixed first order inclusion probabilities

It is usual in practice to search for sampling designs with given first order inclusion probabilities $\pi_k = \Pi_k$, where Π is a vector of size N such that $0 \leq \Pi_k \leq 1$. Among them, statisticians are particularly interested in fixed-size sampling designs, equal-weighted sampling designs (π_k is constant for all k), or both. We first prove that fixed size determinantal sampling designs with fixed first order inclusion probabilities always exist. Then we describe a construction of fixed size equal-weighted determinantal sampling designs. Finally, determinantal sampling designs with the same first and second order inclusion probabilities as SRS are studied.

2.1. Fixed-size determinantal sampling designs with fixed first order inclusion probabilities

Let Π be a vector of size N such that $0 \leq \Pi_k \leq 1$. There exists a very simple determinantal sampling design satisfying $\pi_k = \Pi_k$ for all k : the Poisson sampling (1.5) with kernel K defined by $K_{kk} = \Pi_k$ and $K_{kl} = 0, k \neq l$. Unfortunately this design is not of fixed size. Next theorem proves that fixed-size determinantal sampling designs with fixed first order inclusion probabilities exist under the sole assumption that $\sum_{k=1}^N \Pi_k$ be an integer.

THEOREM 4. *Let Π be a vector of size N such that $0 \leq \Pi_k \leq 1$ and $\sum_{k=1}^N \Pi_k = n \in \mathbb{N}$. There exists a determinantal sampling design of fixed size n whose first order inclusion probabilities satisfy $\pi_k = \Pi_k$.*

PROOF. We use the Schur-Horn Theorem (Horn (1954)). Set $\sum_{k=1}^N \Pi_k = n, \lambda_k = 1$ for $1 \leq k \leq n$ and $\lambda_k = 0$ for $n+1 \leq k \leq N$. Then for all $1 \leq l \leq N$

$$\sum_{i=1}^l \Pi_i \leq l \wedge n \leq \sum_{k=1}^l \lambda_k$$

and

$$\sum_{k=1}^N \Pi_k = n = \sum_{k=1}^N \lambda_k.$$

From the Schur-Horn Theorem, there exists an Hermitian matrix K whose eigenvalues are $\{\lambda_k, 1 \leq k \leq N\}$ and with diagonal elements $\{\Pi_k, 1 \leq k \leq N\}$. As K is a projection with rank n , the associated determinantal design \mathcal{P}_K is of fixed-size n .

Kadison (2002) proposes an algorithm based on rotations to construct such a projection. More generally, Dhillon (2005) describes algorithms to construct Hermitian matrices of given diagonal and spectrum.

2.2. Fixed-size equal-weighted sampling designs

Sampling designs with constant first order inclusion probabilities are a particular instance of sampling designs with fixed first order inclusion probabilities, called equal-weighted sampling designs. By Theorem 4, there exists a fixed-size determinantal sampling designs with fixed first order inclusion probabilities equal to c whenever cN is an integer n . Next theorem gives an explicit formula of the kernel of such determinantal sampling designs, based on the primitive N th roots of unity.

THEOREM 5. *Let n, r, N be three integers such that $n \leq N$ and $r < N$ with r, N relatively prime. let $\mathcal{P}^{r, N, n}$ be the determinantal sampling design with kernel $P^{r, N, n}$:*

$$\begin{cases} P_{kl}^{r, N, n} = \frac{1}{N} \frac{\sin(\frac{nr(k-l)\pi}{N})}{\sin(\frac{r(k-l)\pi}{N})} e^{\frac{ir(n-1)(k-l)\pi}{N}}, \\ P_{kk}^{r, N, n} = \frac{n}{N}. \end{cases}$$

Sampling design $\mathcal{P}^{r, N, n}$ is of fixed size n , and its first and second order inclusion probabilities satisfy

$$\begin{cases} \pi_k^{r, N, n} &= \frac{n}{N}, \\ \pi_{kl}^{r, N, n} &= \frac{n^2}{N^2} - \frac{1}{N^2} \frac{\sin^2(\frac{nr(k-l)\pi}{N})}{\sin^2(\frac{r(k-l)\pi}{N})} \quad (k \neq l). \end{cases}$$

Proof Let $n \leq N$ and $r < N$ with r, N relatively prime be fixed integers. The polynomial $z^N - 1$ admits N roots (in \mathbb{C}) and $\phi(N)$ primitive roots, where ϕ denotes Euler's totient function. Recall that a N th root of unity z is primitive iff N is the smallest integer d such that $z^d = 1$, iff the group $\{z, \dots, z^N = 1\}$ is cyclic of order N . Let $z = e^{\frac{2i\pi r}{N}}$ be a any primitive N th root. Set $c = n/N$ and define for all $p = 0, \dots, n-1$ the vectors

$$\overline{V}_p^T = \frac{\sqrt{c}}{\sqrt{n}} ((z^p)^1, \dots, (z^p)^N),$$

Set $p \in \{0, \dots, n-1\}$. Then

$$\overline{V_p}^T V_p = \sum_{j=1}^N \overline{V_p}^T(j) V_p(j) = n^{-1} c \sum_{j=1}^N |z|^{2pj} = n^{-1} c N = 1.$$

Set $p, q \in \{0, \dots, n-1\}$ and assume $p > q$. Then

$$\overline{V_p}^T V_q = n^{-1} c \sum_{j=1}^N z^{jp} \overline{z}^{jq} = n^{-1} c \sum_{j=1}^N (z^{p-q})^j = 0.$$

Indeed, for any N th root α of unity distinct from 1,

$$\sum_{j=1}^N \alpha^j = \alpha \sum_{j=0}^{N-1} \alpha^j = (1 - \alpha^N)(1 - \alpha)^{-1} = 0.$$

As z is primitive and $1 \leq p - q \leq N - 1$ then z^{p-l} is distinct from 1 and $\overline{V_p}^T V_q = 0$. It follows that V_1, \dots, V_n is an orthonormal family.

By the previous result, $P^{r,N,n} = \sum_{p=0}^{n-1} V_p \overline{V_p}^T$ is a projection of rank n . Its diagonal elements satisfy $P_{k,k}^{r,N,n} = \sum_{p=0}^{n-1} V_p(k) \overline{V_p}^T(k) = n^{-1} c \sum_{p=0}^{n-1} 1 = c$ for all $k = 1, \dots, N$. Its off-diagonal elements satisfy

$$\begin{aligned} P_{kl}^{r,N,n} &= \frac{1}{N} \sum_{p=0}^{n-1} z^{(k-l)p} = \frac{1}{N} \frac{1 - z^{(k-l)n}}{1 - z^{(k-l)}} \\ &= \frac{1}{N} \frac{1 - e^{\frac{2i\pi r(k-l)n}{N}}}{1 - e^{\frac{2i\pi r(k-l)}{N}}} = \frac{1}{N} \frac{\sin\left(\frac{nr(k-l)\pi}{N}\right)}{\sin\left(\frac{r(k-l)\pi}{N}\right)} e^{\frac{ir(n-1)(k-l)\pi}{N}} \quad (k \neq l), \end{aligned}$$

The second order inclusion probabilities follows from (3). □

Figure 1 shows those probabilities in the following case : $N = 18, n = 6, r \in (1, 5, 7)$.

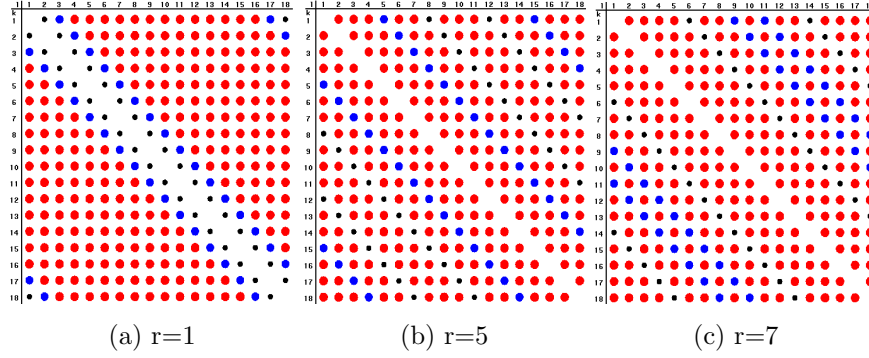
COROLLARY 5.1. *Let n, r, N be three integers such that $n \leq N$ and $r < N$ with r, N relatively prime. Then*

$$\sum_{k=1}^{N-1} \frac{\sin^2\left(\frac{nrk\pi}{N}\right)}{\sin^2\left(\frac{rk\pi}{N}\right)} = n(N - n)$$

Proof We apply the classical result for the fixed-size sampling designs : $\sum_{k \neq l} \pi_{kl} = (n-1)\pi_l$ (see Sarndal (1992) for instance) with $l = N$. □

Fig. 1. Study of $\mathcal{P}^{r,18,6}$ for $r \in (1, 5, 7) : \pi_k^{r,18,6} = \pi_k = \frac{6}{18}$

Circle with radius proportionnal to $\frac{\pi_{kl}^{r,18,6}}{\pi_k \pi_l} (k \neq l)$ ($0 \leq \text{black} \leq 0.5$, $0.5 \leq \text{blue} \leq 0.9$, $0.9 \leq \text{red} \leq 1$.)



2.3. (N, n) -simple sampling designs

We conclude this section with a study of determinantal sampling designs whose first and second order inclusion probabilities are exactly equal to those of SRS. First, we note that SRS is not determinantal in general.

LEMMA 1. *SRS of size n is determinantal iff $n \in \{0, 1, N - 1, N\}$.*

Proof SRS of size $0, 1, N - 1, N$ are associated to the matrices $0, \frac{1}{N}J_N, I_N - \frac{1}{N}J_N$ and I_N respectively. Let $n \notin \{0, 1, N, N - 1\}$, and assume that SRS is a determinantal design. Let $D = \{1, 2\}$ be the domain with the first two elements of the population. Then by Proposition 1.4 the restriction of the design to the domain D is also a determinantal sampling design and by Theorem 1, the number of points ν falling in D has the same law as the sum of two binomial random variables. Let p_1 and p_2 be the respective parameter of these random variables. Then $p_1 + p_2 = \mathbb{E}(\nu)$ and $p_1 p_2 = pr(\nu = 2)$. Computing these values using SRS we get that p_1 and p_2 are the zeros of the second order polynomial equation $n(N - 1)X^2 - 2N(N - 1)X + n(n - 1) = 0$. But one of the solutions is more than $\frac{N}{n} > 1$, and cannot be a probability. \square

This negative result does however not settle the question of the existence of a determinantal sampling design with the same first and second order inclusion probabilities as the SRS of size n , that is such that $\pi_k = \frac{n}{N}$ et $\pi_{kl} = \frac{n(n-1)}{N(N-1)} (k \neq l)$.

DEFINITION 2.1. *A determinantal sampling design \mathcal{P} (on U of size N) is (N, n) -simple if its inclusion probabilities satisfy $\pi_k = \frac{n}{N}$ et $\pi_{kl} = \frac{n(n-1)}{N(N-1)}$.*

Applying (2) and (3) to a (N, n) -simple determinantal sampling design, we get that its kernel K satisfies

$$\begin{aligned} K_{k,k} &= \frac{n}{N}, \\ |K_{kl}|^2 &= \frac{n(N-n)}{N^2(N-1)} (k \neq l); \end{aligned}$$

It follows from these equations and Proposition 1.3 that the complementary sample \mathbb{S}^c of a (N, n) -simple determinantal random sample \mathbb{S} is $(N, N-n)$ -simple.

LEMMA 2. *Let $n \leq N$ be two integers and \mathcal{P} a (N, n) -simple determinantal sampling design. Then \mathcal{P} is of fixed size n .*

Proof Let $\#\mathbb{S}$ be the cardinal of the random sample \mathbb{S} with law \mathcal{P} . Its variance depends only on the first and second order inclusion probabilities of the sampling design \mathcal{P} , which are those of SRS. As SRS is of fixed size, $\text{var}(\#\mathbb{S}) = 0$, i.e. \mathcal{P} is of fixed size. \square

THEOREM 6. *Let $1 < n < N-1$ be two integers. Then there exists a (N, n) -simple determinantal sampling design only if $N \leq \min\{n^2, (N-n)^2\}$.*

Proof Let K be the kernel of such a design. By Lemma 2 and Corollary 1.1, it is a rank n projection. Recall that the Schur-Hadamard product is rank submultiplicative ($\text{rank}(A * B) \leq \text{rank}(A)\text{rank}(B)$). Then

$$K * \overline{K} = \alpha I_N + \beta J_N \tag{9}$$

with $\alpha = -\frac{n(n-1)}{N(N-1)}$ and $\beta = \frac{n(N-n)}{N^2(N-1)}$. Indeed $(K * \overline{K})_{k,k} = \frac{n^2}{N^2}$ for all $1 \leq k \leq N$ and $(K * \overline{K})_{kl} = |K_{kl}|^2 = \frac{n(N-n)}{N^2(N-1)}$ for all $k \neq l$. As the eigenvalues of J_N are 0 and N and $\alpha \notin \{0, -\beta N\}$, $K * \overline{K}$ is invertible hence of rank N . By submultiplicativity of the rank $\text{rank}(K * \overline{K}) \leq \text{rank}(K)\text{rg}(\overline{K}) = \text{rank}^2(K)$ and finally $N \leq n^2$. Symmetrically, the same arguments on the projection matrix $I_N - K$ (associated to the $(N, N-n)$ -simple complementary sampling design \mathcal{P}^c) of rank $N-n$ gives $N \leq (N-n)^2$. \square

The proof of the theorem is adapted from similar results in the context of equiangular frames (Tropp (2005)). Indeed, Equiangular Tight Frames (ETFs) and (N, n) -simple determinantal sampling design are equivalent notions, in a sense made precise below. An ETF of size N on \mathbb{C}^n can be defined as a family of unit vectors $\{f_1, \dots, f_N\}$ such that the absolute values of the inner products between pairs of vectors are both identical and minimal. Equivalently, an ETF is a $n \times N$ matrix F such that:

- (a) For all $l \in 1, \dots, N$, $\sum_{k=1}^n F_{kl}^2 = 1$;
- (b) There exists a nonnegative α such that $|\sum_{j=1}^n \overline{F_{jk}} F_{jl}| = \alpha$ ($k \neq l$);
- (c) $F \overline{F}^T = \frac{N}{n} I_n$.

Table 1. Existence of (N, n) -simple determinantal sampling designs, depending on the kernel type (real or complex) for $n < 9$.

n	3	3	4	4	5	5	6	6	6	7	7	7	8	8	8
N	6	7	7	13	10	11	11	16	31	14	15	28	15	29	57
	\mathbb{R}	\mathbb{C}	\mathbb{C}	\mathbb{C}	\mathbb{R}	\mathbb{C}	\mathbb{C}	\mathbb{R}	\mathbb{C}	\mathbb{R}	\mathbb{C}	\mathbb{R}	\mathbb{C}	\mathbb{C}	\mathbb{C}

A non obvious consequence of this definition is that the “angle” α is fixed and equal to $\sqrt{\frac{N-n}{n(N-1)}}$ (Sustik (2007) Proposition 2). An other consequence is that the eigenvalues of $\bar{F}^T F$ are $\frac{N}{n}$ and 0 (unless $n = N$). It follows that matrix $K = \frac{n}{N} \bar{F}^T F$ is the kernel of a (N, n) -simple determinantal sampling design. Conversely, if K is such a kernel then it follows from the spectral decomposition $K = P\Gamma\bar{P}^T$ with $\Gamma_{kk} \in \{0, 1\}$ that the eigenvectors associated to 1 scaled by a factor $\sqrt{\frac{N}{n}}$ define a ETF F . We have shown that:

THEOREM 7. *Matrix K is the kernel of a (N, n) -simple determinantal sampling design if and only if there exists an ETF F of size N on \mathbb{C}^n such that $K = \frac{n}{N} \bar{F}^T F$.*

The matrices in Theorem 6 are assumed to be complex. If one considers real matrices only, then more restrictive conditions are needed. For instance the bound in Theorem 6 is $N \leq \min\{\frac{n(n+1)}{2}, \frac{(N-n)(N-n+1)}{2}\}$ (Sustik (2007) Theorem C). From Sustik (2007) Theorem A and Casazza (2008) Theorem 4.1 we also deduce :

THEOREM 8. *Let $1 < n < N - 1$. When $N \neq 2n$, a necessary condition of the existence of a (N, n) -simple determinantal sampling design with real kernel K is that the following two quantities be odd integers:*

$$\alpha = \sqrt{\frac{n(N-1)}{N-n}}, \quad \beta = \sqrt{\frac{(N-n)(N-1)}{n}}.$$

When $N = 2n$, it is necessary that n be odd and that $N - 1$ be the sum of two squares.

As a consequence of Theorem 7, a necessary and sufficient condition for the existence of ETFs would solve the problem of the existence of (N, n) -simple determinantal sampling designs. However, such a condition is not known for the moment. But numerical studies exist. We deduce from Sustik (2007) and Casazza (2008) the existence of (N, n) -simple determinantal sampling designs with real kernel for respectively $N \leq 100$ (Sustik (2007) Table I) and $n \leq 50$ (Casazza (2008) Table III). Tables II and III in Sustik (2007) also give the existence of (N, n) -simple determinantal sampling designs with complex kernels. Table 1 summarizes this information for $n < 9$. In the table, the symbol \mathbb{C} indicates that no (N, n) -simple determinantal sampling design with real kernel exists, but that one with complex kernel does exist.

EXAMPLE 2.1 ((6,3)-SIMPLE DETERMINANTAL SAMPLING DESIGN). *Let*

$$K = \frac{1}{2} \begin{pmatrix} 1 & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & 1 & -\frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 1 & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 1 & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 1 & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 1 \end{pmatrix}.$$

K is a projection, and the associated determinantal sampling design is (6,3)-simple. It is not a simple sampling as the samples $\{1, 2, 3\}$ and $\{4, 5, 6\}$ have not the same probabilities ($\frac{1}{8}(1 - \frac{3}{5} - \frac{2}{5\sqrt{5}})$ and $\frac{1}{8}(1 - \frac{3}{5} + \frac{2}{5\sqrt{5}})$ respectively).

We note that some (non-determinantal) SRS have determinantal analogues (that is with the same first and second order inclusion probabilities). In the general case, define, for any sampling design \mathcal{P} and any complex matrix ζ , the matrix $K^{\Delta, \zeta} = \zeta * K^{\Delta}$ where

$$K_{kl}^{\Delta} = \begin{cases} \sqrt{\pi_{kl} - \pi_k \pi_l} & (k \neq l) \\ \pi_k & (k = l) \end{cases}$$

The square root is taken in the complex sense for sampling designs that do not verify the Sen-Yates-Grundy condition (6). Then there exists a determinantal sampling design with the same first and second order inclusion probabilities as \mathcal{P} iff there exists a matrix $\zeta^{\mathcal{P}}$ such that $\zeta_{kl}^{\mathcal{P}} = \bar{\zeta}_{lk}^{\mathcal{P}} = \frac{1}{\zeta_{kl}^{\mathcal{P}}}$ ($k \neq l$), $\zeta_{kk}^{\mathcal{P}} = 1$ and $K^{\Delta, \zeta^{\mathcal{P}}}$ is a contracting matrix.

In the previous example, the matrix $\zeta^{\mathcal{P}}$ was

$$\zeta^{\mathcal{P}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}.$$

3. Estimation of a total

3.1. Mean Square Error

In the case of a determinantal sampling design, the MSE of an homogeneous linear estimator of the total t_y of a variable of interest y , can be computed using formulas (1) et (4) and rewritten in an algebraic form. This algebraic form itself leads to a geometric formulation of the MSE. This rewriting enables us to provide a necessary and sufficient condition for an exact estimation of the total of auxiliary variables.

In this section we consider a variable of interest y , a determinantal sampling design \mathcal{P}_K with kernel K and the homogeneous linear estimator \hat{t}_{yw} of t_y with weights w . We also introduce the following notations. For any vector x , D_x denotes the diagonal matrix with diagonal x . For any two matrices $A, B \in \mathcal{M}_N(\mathbb{C})$, $\langle A, B \rangle = \text{tr}(\overline{A}^T B) = \sum_{k,l} \overline{a_{k,l}} b_{k,l}$ denotes the canonical scalar product on $\mathcal{M}_N(\mathbb{C})$. The associated Frobenius norm is denoted by $|A|$. We also define $z = w * y$ and diagonal matrix Z by $Z_{kk} = \sqrt{w_k y_k}$ ($Z = D_{\sqrt{w*y}}$) where the square root is taken in the complex sense for negative y . Finally, we pose $\langle\langle A, B \rangle\rangle = \langle \overline{Z}^T A Z, Z B \overline{Z}^T \rangle$. Note that $Z = Z^T$ and $\overline{Z}^2 = Z^2$, equalities that we will use thoroughly in the sequel.

PROPOSITION 3.1 (ALGEBRAIC AND GEOMETRIC FORMS OF THE MSE). *The MSE of \hat{t}_{yw} satisfies*

$$\text{MSE}(\hat{t}_{yw}) = z^T((I_N - K) * \overline{K})z + [e^T(K * I_N)z - e^T y]^2 \quad (10)$$

$$= \langle \overline{Z}^T (I_N - K) Z, Z K \overline{Z}^T \rangle + [\langle D_w K - I_N, D_y \rangle]^2 \quad (11)$$

$$= \langle\langle I_N - K, K \rangle\rangle + [\langle D_y, K D_w - I_N \rangle]^2 \quad (12)$$

Proof These formulas follow from the classical equality $\text{tr}(AB) = \text{tr}(BA)$ and the following equality relating the trace on the Schur-Hadamard product (Horn (1991)): for any two vectors x, y and any two matrices A, B it holds that

$$\overline{x}^T A * B y = \text{tr}(\overline{D_x} A D_y B^T).$$

We then have

$$\begin{aligned} \text{MSE}(\hat{t}_{yw}) &= z^T((I_N - K) * \overline{K})z + [e^T(K * I_N)z - e^T y]^2 \\ &= \text{tr}(\overline{D_z}(I_N - K)D_z \overline{K}^T) + [\text{tr}(\overline{D_e} K D_z I_N^T) - \text{tr}(D_y)]^2 \\ &= \text{tr}(\overline{Z}^2(I_N - K)Z^2 K) + [\text{tr}(K D_w D_y - D_y)]^2 \\ &= \text{tr}(\overline{Z}(I_N - K)Z Z K \overline{Z}) + [\text{tr}(D_y(K D_w D_y - I_N))]^2 \\ &= \langle \overline{Z}^T (I_N - K) Z, Z K \overline{Z}^T \rangle + [\langle D_y, K D_w - I_N \rangle]^2 \end{aligned}$$

□

The bilinear form $\langle\langle \cdot, \cdot \rangle\rangle$ is indefinite in general. However, it holds by Moutard-Fejer's Theorem (de Klerk (2006) Appendix A) that for any two positive semidefinite matrices A and B

$$\langle\langle A, B \rangle\rangle = \langle \overline{Z}^T A Z, Z B \overline{Z}^T \rangle \geq 0$$

since $\overline{Z}^T A Z$ and $Z B \overline{Z}^T$ are positive semidefinite.

For the Horvitz-Thompson estimator we get:

PROPOSITION 3.2.

$$\text{MSE}(\hat{t}_{yHT}) = \text{var}(\hat{t}_{yHT}) = (\pi^{-1} * y)^T((I_N - K) * \overline{K})(\pi^{-1} * y)$$

or geometrically

$$\text{MSE}(\hat{t}_{yHT}) = \text{var}(\hat{t}_{yHT}) = \langle \bar{Z}^T (I_N - K) Z, Z K \bar{Z}^T \rangle = \langle \langle I_N - K, K \rangle \rangle$$

Recently, Deville (2014) raised the following question. For a given vector y , when can we estimate perfectly (without error, $\text{MSE} = 0$) the total y , using a sampling design with fixed first order inclusion probabilities (and an homogeneous linear estimator)? Using the previous equations, we provide a necessary and sufficient condition within determinantal sampling designs. Obviously, the estimator must be unbiased. Therefore we consider the Horvitz-Thompson estimator only ($w_k = \frac{1}{\pi_k}$, $k = 1, \dots, N$), and positive first order inclusion probabilities.

PROPOSITION 3.3. *Assume y takes only non-zero values. Then the total t_y will be perfectly estimated ($\text{MSE} = 0$) by the Horvitz-Thompson estimator associated to \mathcal{P}_K iff K is a projection with positive diagonal that commutes with Z^2 .*

Proof By Moutard-Fejer's Theorem, it holds that for any two semidefinite matrices A and B , $\text{tr}(AB) \geq 0$ with equality iff $AB = 0$. Assume K be a contracting matrix such that $\text{MSE}(\hat{t}_{yHT}) = 0$. Then $\text{tr}(\bar{Z}^T (I_N - K) Z Z K \bar{Z}^T) = 0$. As $\bar{Z}^T (I_N - K) Z$ and $Z K \bar{Z}^T$ are semidefinite, then $\bar{Z}^T (I_N - K) Z^2 K \bar{Z}^T = 0$. Multiplying on the left and on the right by \bar{Z}^{-1} yields $Z^2 K = K Z^2 K$ and taking the conjugate transpose gives $Z^2 K = K Z^2 K = K Z^2$. It then follows that $Z^2 K = Z^2 K^2$. By multiplying the equality on the left by Z^{-2} we get $K^2 = K$, that is K is a projection. Conversely, if K is a projection that commutes with Z^2 then $\bar{Z}^T (I_N - K) Z^2 K \bar{Z}^T = \bar{Z}^T Z^2 (I_N - K) K Z = 0$ and by taking the trace, $\text{MSE}(\hat{t}_{yHT}) = 0$. \square

The proof can be adapted to the case y takes the value 0. Using the Moore-Penrose inverse Z^\dagger of Z (that is also its group inverse, $ZZ^\dagger = Z^\dagger Z$, Ben-Israel (2003)) and Peirce decomposition (Peirce (1881), Lam (2001)), one obtains as a necessary and sufficient condition “ $P = Z^\dagger Z K Z^\dagger Z$ is a projection that commutes with Z^2 ”.

The existence of such a matrix is highly constrained, notably when the first order inclusion probabilities are fixed. Indeed, it is a classical exercise to show that the commutant of a diagonal matrix is a certain set of block diagonal matrices. Let Π be a probability vector and $\alpha_1, \dots, \alpha_q$ be the distinct values of $\{\frac{y_k}{\Pi_k}, k = 1, \dots, N\}$, and $A_j, j = 1, \dots, q$ be the associated sets of indices k such that $\frac{y_k}{\Pi_k} = \alpha_j$.

COROLLARY 8.1. *With the previous notation, there exists a determinantal sampling design \mathcal{P} with first order inclusion probabilities $\pi_k = \Pi_k$ such that the associated Horvitz-Thompson estimator \hat{t}_{yHT} estimates perfectly y iff*

$$\forall j = 1, \dots, q, \sum_{k \in A_j} \pi_k \text{ is a positive integer.}$$

In that case, the determinantal sampling designs that solve the problem are the stratified determinantal sampling designs with strata A_j , $j = 1, \dots, q$, that respect the first order inclusion probabilities and are of fixed size within each strata.

3.2. Statistical properties of the estimator

The classical settings for the study of asymptotic properties are either the superpopulation models (Deming (1941), Cassel (1977) chapter 4), or the models of nested (finite) populations as described by Isaki and Fuller (1982). We consider this second setting here. In particular, $(U_N, N \in \mathbb{N})$ is a nested sequence of finite populations ($U_N \subseteq U_{N+1}$). To simplify notation, we identify these finite populations with the sets $\{1, \dots, r_N\}$, where $r_N = \#U_N$ (in practice $r_N = N$ in general). The variable of interest y is a sequence of values $(y_k, k \in \mathbb{N})$. In this section we focus successively on consistency, central limit theorems and concentration/deviation inequalities for linear estimators of the total. We close the section by considering the finite distance law of the estimator of the population size of a domain.

In this setting, most results about consistency concern the mean square convergence of the Horvitz-Thompson estimator of the mean (see Isaki and Fuller (1982), Robinson (1982), Dol (1996) in the case of fixed-size sampling designs and Cardot (2010), Chauvet (2014) in the general case). A classical condition within these reference is that the sequence $\frac{1}{r_N} \sum_{k \in U_N} \frac{y_k^2}{(\pi_k^N)^2}$ is bounded. Using Schur's Theorem (Schur (1911)) on semidefinite matrices we improve the previous condition for determinantal sampling designs. The theorem also applies to other linear homogeneous estimators than the Horvitz-Thompson one.

THEOREM 9. *Let $(\mathcal{P}_N, N \in \mathbb{N})$ be a sequence of determinantal sampling designs on the populations $U_N = \{1, \dots, r_N\}$ with kernel $(K^N, N \in \mathbb{N})$, whose diagonal terms are positive, and $(w^N, N \in \mathbb{N})$ a sequence of positive vectors of size r_N . If*

$$(a) \sum_{k \in U_N} K_{kk}^N \left(1 - \frac{1}{K_{kk}^N w_k^N}\right)^2 = O(1),$$

$$(b) \frac{1}{r_N^2} \sum_{k \in U_N} K_{kk}^N (w_k^N y_k)^2 \xrightarrow{N \rightarrow \infty} 0,$$

then $\frac{\hat{t}_{yw}^N - t_y^N}{r_N}$ tends to 0 in mean square.

In particular a sufficient condition for the convergence of $\frac{\hat{t}_{yHT}^N - t_y^N}{r_N}$ towards 0 in mean square is

$$\frac{1}{r_N^2} \sum_{k \in U_N} \frac{y_k^2}{K_{kk}^N} \xrightarrow{N \rightarrow \infty} 0.$$

Proof Let \mathcal{P}_K be a determinantal sampling design with kernel K , π its vector of diagonal terms (first order inclusion probabilities). By Proposition 3.1

$$\text{MSE}(\hat{t}_{yw}) = (w * y)^T ((I_N - K) * \bar{K})(w * y) + [e^T (I_N * K)(w * y) - e^T y]^2$$

As the matrices I , K , $(I - K)$ and \bar{K} are positive semidefinite, it holds that $((I - K) * \bar{K})$, $I * \bar{K}$ and $K * \bar{K}$ are positive semidefinite by Schur Theorem. As also $(I - K) * \bar{K} = I * \bar{K} - K * \bar{K}$ then $(I - K) * \bar{K} \leq I * \bar{K}$ for the

partial order on positive semidefinite matrices. It follows that

$$\begin{aligned} (w * y)^T (I - K) * \bar{K} (w * y) &\leq (w * y)^T (I * \bar{K}) (w * y) \\ &\leq \sum_{k \in U} (w_k y_k)^2 K_{kk}; \end{aligned}$$

Also the bias satisfies

$$\begin{aligned} [e^T (I_N * K) (w * y) - e^T y]^2 &= \left(\sum_{k \in U} (K_{kk} - \frac{1}{w_k}) (w_k y_k) \right)^2 \\ &= \left(\sum_{k \in U} (\sqrt{K_{kk}} - \frac{1}{\sqrt{K_{kk} w_k}}) (\sqrt{K_{kk} w_k} y_k) \right)^2 \\ &\leq \left(\sum_{k \in U} (\sqrt{K_{kk}} - \frac{1}{\sqrt{K_{kk} w_k}})^2 \right) \left(\sum_{k \in U} (K_{kk} (w_k y_k)^2) \right) \end{aligned}$$

by Cauchy-Schwartz-inequality. From these inequations we get

$$E \left(\left(\frac{\hat{t}_{yw}^N - t_y^N}{r_N} \right)^2 \right) \leq \left(1 + \sum_{k \in U_N} K_{kk}^N \left(1 - \frac{1}{K_{kk}^N w_k^N} \right)^2 \right) \frac{1}{r_N^2} \left(\sum_{k \in U_N} K_{kk}^N (w_k^N y_k)^2 \right)$$

which gets to 0 by assumptions. This completes the proof. \square

Regarding equal-weighted determinantal sampling designs with expected size μ_N and a bounded variable y , a sufficient condition for convergence of the Horvitz-thompson estimator of the mean is simply $\mu_N \rightarrow \infty$. More generally

COROLLARY 9.1. *Let $(\mathcal{P}_N, N \in \mathbb{N})$ be a sequence of determinantal sampling designs on the populations $U_N = \{1, \dots, r_N\}$ with kernel $(K^N, N \in \mathbb{N})$ of trace μ_N . Assume moreover that*

- (a) *there exists $c > 0$, for all $N \in \mathbb{N}$ and $k \leq N$ $c \frac{\mu_N}{r_N} \leq K_{kk}^N$;*
- (b) *the sequence $(\frac{1}{r_N} \sum_{k \in U_N} y_k^2, N \in \mathbb{N})$ is bounded;*
- (c) *the expected size of the samples $\mu_N \rightarrow \infty$.*

Then $\frac{\hat{t}_{yHT}^N - t_y^N}{r_N} \rightarrow 0$ in mean square.

The second assumption appears for instance in Robinson (1982).

Apart consistency, some authors have considered the existence of central limit theorems for sampling designs. Their results either focus on a particular class of sampling designs (simple random sampling without replacement

: Erdos (1959), Hajek (1960), rejective Poisson sampling : Hajek (1960)), or assume entropy conditions (Berger (1998)). Assuming only that the determinantal sampling design is “random enough”, we obtain a central limit theorem by applying the results of Soshnikov (Soshnikov (2000), Soshnikov (2002)). These articles contain many theorems on the asymptotic normality of functionals of determinantal point processes. Theorem 1 on linear statistics of bounded measurable functions in Soshnikov (2002) can be applied straightforwardly to the study of determinantal sampling designs and the associated linear homogeneous estimators (whose weight do not depend on the random sample).

THEOREM 10. *Let $(\mathcal{P}_N, N \in \mathbb{N})$ be a sequence of determinantal sampling designs on $U_N = \{1, \dots, r_N\}$ with kernels $(K^N, N \in \mathbb{N})$. Let also $(w^N, N \in \mathbb{N})$ be a sequence of positive vectors of size r_N and define for all $N \in \mathbb{N}$ the homogeneous linear estimators*

$$\hat{t}_{yw}^N = \sum_{k \in \mathbb{S}_N} w_k y_k \text{ and } \hat{t}_{|y|w}^N = \sum_{k \in \mathbb{S}_N} w_k |y_k|$$

If the variance $\text{var}(\hat{t}_{yw}^N) \rightarrow +\infty$ as $N \rightarrow \infty$ and if,

$$\sup |w_k y_k| = o(\text{var}(\hat{t}_{yw}^N))^\epsilon \text{ and } E(\hat{t}_{|y|w}^N) = O(\text{var}(\hat{t}_{yw}^N))^\delta$$

for any $\epsilon > 0$ and some $\delta > 0$ then

$$\frac{\hat{t}_{yw}^N - E(\hat{t}_{yw}^N)}{\sqrt{\text{var}(\hat{t}_{yw}^N)}} \xrightarrow{\text{law}} \mathcal{N}(0, 1).$$

As usual, we can replace the true variance $\text{var}(\hat{t}_{yw}^N)$ by any weakly consistent estimator of this variance using Slutsky theorem. A classical variance estimator is the Horvitz-Thompson estimator of the variance (Horvitz (1952)):

$$\widehat{\text{var}}_{HT}(\hat{t}_{yw}) = \sum_{k \in \mathbb{S}} \sum_{l \in \mathbb{S}} \frac{w_k y_k w_l y_l}{\pi_{kl}} \Delta_{kl}$$

In case of fixed-size sampling designs, an alternative formula for the variance can be used and we get the Sen-Yates-Grundy estimator (Yates (1953), Sen (1953)):

$$\widehat{\text{var}}_{SYG}(\hat{t}_{yw}) = \frac{1}{2} \sum_{k \in \mathbb{S}} \sum_{l \in \mathbb{S}} \frac{(w_k y_k - w_l y_l)^2}{\pi_{kl}} \Delta_{kl}$$

which is itself a Horvitz-Thomson estimator (but of a different sum). Unfortunately, the law of the random sample \mathbb{S}' defined on $U \times U$ by $(k, l) \in \mathbb{S}'$ iff $k, l \in \mathbb{S}$ is not determinantal and Theorem 9 does not apply. The first and second order inclusion probabilities of this new sampling design \mathcal{P}' can however be easily calculated by means of the matrix K associated to \mathbb{S} , and we can use the classical criteria of convergence of the Horvitz-Thompson estimator that depend only on these inclusion probabilities (Isaki and Fuller (1982), Robinson (1982), Dol (1996), Cardot (2010), Chauvet (2014)).

More recently, the asymptotic normality has been extended to more general classes of processes (that include the determinantal ones): the processes with negative or positive associations (Yuan (2003)), and the processes that satisfy the strong Rayleigh property (Brandèn (2012)). As previously recorded, from a very different perspective, the work of Berger (1998) proves asymptotic normality under asymptotically maximal entropy conditions.

For processes satisfying the strong Rayleigh property, Pemantle and Peres (2014) recently proved concentration and deviation inequalities. Their application to sampling theory allows to derive the following finite distance results.

THEOREM 11. (*Deviation and concentration inequalities*) *Let \mathcal{P}_K be a determinantal sampling design with kernel K of trace μ . Let y be a variable of interest and \hat{t}_{yw} a linear homogeneous estimator of the total t_y of y whose weights do not depend on the random sample \mathbb{S} . We note $C = \max_{k \in U} |w_k y_k|$. For all $a > 0$,*

$$\begin{aligned} \text{pr}(\hat{t}_{yw} - E(\hat{t}_{yw}) > a) &\leq 3 \exp\left(-\frac{a^2}{16(aC + 2\mu C^2)}\right), \\ \text{pr}(|\hat{t}_{yw} - E(\hat{t}_{yw})| > a) &\leq 5 \exp\left(-\frac{a^2}{16^2(aC + 2\mu C^2)}\right). \end{aligned}$$

Moreover, if \mathcal{P}_K is of fixed size $\mu = n$, then

$$\begin{aligned} \text{pr}(\hat{t}_{yw} - E(\hat{t}_{yw}) > a) &\leq \exp\left(-\frac{a^2}{8nC^2}\right), \\ \text{pr}(|\hat{t}_{yw} - E(\hat{t}_{yw})| > a) &\leq 2 \exp\left(-\frac{a^2}{8nC^2}\right). \end{aligned}$$

Proof Fonction $s \mapsto \sum_{k \in U} w_k y_k 1_{\{k \in s\}}$ is C -Lipschitz for the Hamming distance. Theorems 3.1 and 3.2 of Pemantle and Peres (2014) apply and yield the desired results. \square

From this concentration inequality, we also derive a new criterion for the convergence in probability of \hat{t}_{yHT} :

COROLLARY 11.1. *Let $(\mathcal{P}_N, N \in \mathbb{N})$ be a sequence of determinantal sampling designs on $U_N = \{1, \dots, r_N\}$ with kernel $\{K_N, n \in \mathbb{N}\}$ whose trace $\mu_N = \sum_{k \in U_N} K_N(k, k) > c > 0$. If*

$$\frac{\sqrt{\mu_N}}{r_N} \max_{k \in U_N} \left| \frac{y_k}{K_N(k, k)} \right| \xrightarrow{N \rightarrow \infty} 0$$

then

$$\frac{(\hat{t}_{yHT} - t_y)}{r_N} \xrightarrow[N \rightarrow \infty]{\text{pr}} 0.$$

PROOF. Set $C_N = \max_{k \in U_N} \frac{|y_k|}{K_N(k, k)}$. It holds that

$$\text{pr}(|\hat{t}_{yHT} - t_y| > r_N a) \leq 5 \exp\left(-\frac{r_N^2 a^2}{16^2(r_N a C_N + 2\mu_n C_N^2)}\right)$$

But by assumption $C_N = o(r_N)$ and $2\mu_n C_N^2 = o(r_N)^2$, and the right hand term above tends to 0.

3.3. Estimation of the population size of a domain

Consider the problem of estimating the population size of a given domain D by an equal weight determinantal sampling design. Then the finite distance laws of the estimator (conditionnaly on D) are known. Indeed let $D \subseteq U$ be a domain of population $U = \{1, 2, \dots, N\}$, and define the indicator variable $y_k = 1_{k \in D}$. Estimate $\theta = \#D$ using an equal weight determinantal sampling design with kernel K and diagonal terms $K_{k,k} = \pi_k = c$. Then $\theta = \sum_{k=1}^N y_k$ and the Horvitz-Thomson estimator can be rewritten $\hat{\theta} = c^{-1} \sum_{k \in \mathbb{S}} y_k = c^{-1} \#(D \cap \mathbb{S})$. A direct application of Theorem 1 then gives:

COROLLARY 11.2. *The random variable $c\hat{\theta}$ has Poisson-binomial law, the law of the sum of $\#D$ independant Bernoulli random variables with parameter λ_i , eigenvalues of the matrix $K|_D$.*

PROOF. By Proposition 1.4, the restriction of the random determinantal sample \mathbb{S} to the domain D is a random determinantal sample with kernel $K|_D$. By Theorem 1, the number of points of this sample has Poisson-binomial law.

The Lindeberg-Feller theorem for triangular arrays applied to the sum of independant Bernoulli random variables (with possibly distinct parameters) gives the previous Central Limit Theorem. Berry-Essen type theorems then gives, additionnaly to the asymptotic normality, a convergence rate. Finally, Bernstein inequality provides a deviation inequality.

4. Optimal determinantal sampling designs

Within a given family of sampling designs, it is common to search for optimal designs with respect to criteria that depend on the context and the wishes of the statistician. For instance, Neyman (1934) found the stratified sampling design that minimizes the variance of the Horvitz-Thompson estimator \hat{t}_{yHT} where y is a given variable. In the same spirit, Chauvet, Bonnery and Deville (2011) minimize the approximated variance of \hat{t}_{yHT} within the class of sampling designs balanced on given auxiliary variables.

In this section, we adress the problem of finding the best representative sampling design within the parametric class of determinantal sampling designs. We notably show that even in the case of minimizing the MSE, challenging issues may be encountered.

4.1. Representative sampling design strategy

Hajek (1960) defines a sampling design strategy as a couple (\mathcal{P}, w) where \mathcal{P} is a sampling design and w a system of weights. The strategy is called *representative* for a set of Q auxilliary variables $x_q, q = 1, \dots, Q$ if the associated homogeneous linear estimators gives a perfect estimation of the total of each variable: $\text{MSE}(\hat{t}_{x_q w}) = 0$ for all $q = 1, \dots, Q$.

The two components of a strategy are usually not computed simultaneously. For a given vector of first inclusion probabilities, Deville and Tillé (2004) have defined an algorithm of *balanced sampling*, that defines an approximately representative strategy, when the weights are those of the Horvitz-Thompson estimator: $\text{MSE}(\hat{t}_{x^q HT}) \approx 0$ for all $q = 1, \dots, Q$. Conversely, for a given sampling design, Deville and Sarndal (1992) find a system of weights $w(\mathbb{S})$ that depends on the random sample \mathbb{S} ensuring representativeness, known as calibration.

Regarding determinantal sampling designs, we can conduct the search of both the sampling design and the weights simultaneously, together with penalization terms, by solving an optimization problem of the following type:

$$\underset{(K,w) \in \Theta}{\text{Min}} g(X, K, w) + \lambda \text{pen}(K, w) \quad (13)$$

where g satisfies the balancing equations $g(X, K, w) = 0$ iff (\mathcal{P}_k, w) is representative for X . Function pen is a penalization term to avoid overfitting. This penalization function can also be used to transform a constrained optimization problem into a free optimization problem, and λ is then an adjustment parameter. The unknown parameters are the matrix K and the vector of weights w .

In the following, we list some objective functions, parameter sets and penalization functions that could be interesting to the statistician.

(a) Examples of objective functions g :

- A simple function that characterizes representativity is

$$g(X, K, w) = \sum_{q=1}^Q \text{MSE}(\hat{t}_{x^q w}) \quad (14)$$

- Function

$$g(X, K, w) = \sum_{q=1}^Q \alpha_q \frac{\sqrt{V(\hat{t}_{x^q w})}}{t_{x^q}} + \beta_q \frac{|\text{bias}(\hat{t}_{x^q w})|}{t_{x^q}}$$

offers a tradeoff between bias and variance through parameter β . Parameter α allows an adjustment between the different auxiliary variables. These parameters are chosen by the statistician.

(b) Examples of parameter sets Θ :

- To consider only determinantal sampling designs of fixed size n , we take Θ as the set of projections matrices of rank n , and $w \in \mathbb{R}_+^N$.
- If the size is random with fixed expectation, we use contracting matrices K of fixed trace.
- If we are only interested in the Horvitz-Thompson estimator, we take $\Theta = \{(K, w) | \text{spec}(K) \in [0, 1]^N, w_k = \frac{1}{K_{kk}} (k = 1, \dots, N)\}$

- Let $\Pi \in [0, 1]^N$ be a given vector. Then $\Theta = \{(K, w) | \text{spec}(K) \in [0, 1]^N, K_{kk} = \Pi_k \ (k = 1, \dots, N), w \in \mathbb{R}_+^N\}$ corresponds to the set of determinantal sampling designs with first order inclusion probabilities $\pi_k = \Pi_k, \ (k = 1, \dots, N)$.

(c) Examples of penalization functions pen :

- $pen(K, w) = \text{tr}(K)^d$ penalizes large size sampling designs.
- $pen(K, w) = \sum_k K_{kk}(w_k - \frac{1}{K_{kk}})^2$, convex relaxation (in w) of the constrained set $\Theta = \{(K, w) | w_k = \frac{1}{K_{kk}}, \forall 1 \leq k \leq N\}$, can limit the overfitting effect due to the search of representativity, by analogy with the classical calibration method. Other distance measures can be used (Deville and Sarndal (1992)).
- $pen(K, w) = (\prod_k 1_{0.25 \leq K_{kk}w_k \leq 1.75})^{-2}$ corresponds directly to the constraint $0.25 \leq K_{kk}w_k \leq 1.75$, such that the estimator is “not too far” from the Horvitz-Thompson estimator.
- Let C be the contiguity matrix such that $c_{kl} = 1$ if k and l are contiguous, and 0 otherwise. Then $pen(K, w) = \sum_k \sum_{l > k} c_{kl}(K_{kk}K_{ll} - |K_{kl}|^2)$ penalizes the sampling designs regarding the average number of neighbours in the random sample.

These optimization problems are generally well-posed, because we consider continuous functions on the convex compact set of contracting matrices. These problems are nonetheless non-convex problems, because the objective functions are non-convex. Algorithmic difficulties can be challenging, as shown below.

4.2. A case study: minimization of the Mean Square Error within fixed first order inclusion probabilities determinantal sampling designs

Let U be a population of size N , y be a fixed variable and Π be a fixed vector with values in $[0, 1]$. To simplify the problem, we also consider a fixed system of weights w , and pose $z = w * y$. We want to minimize the MSE over the set of determinantal sampling designs with first order inclusion probabilities $\pi_k = \Pi_k, \ k = 1, \dots, N$, that is over the set $\Theta = \{(K, w) | \text{spec}(K) \in [0, 1]^N, K_{kk} = \Pi_k, \forall 1 \leq k \leq N\}$ of contracting matrices of diagonal Π . This problem admits the following algebraic formulation:

$$(P) = \arg \min_{0 \leq K \leq I_N, \text{diag}(K) = \Pi} z^T((I_N - K) * \overline{K})z + [e^T(I_N * K)z - e^T y]^2.$$

Since the bias $e^T(I_N * K)z - e^T y$ is constant, it holds that

$$(P) = \arg \min_{0 \leq K \leq I_N, \text{diag}(K) = \Pi} z^T((I_N - K) * \overline{K})z.$$

We again interpret this problem geometrically. Let S_N^+ be the convex cone of positive semidefinite matrices. The set of contracting matrices is $S_N^+ \cap (I_N - S_N^+)$ (and $\frac{1}{2}I$ can be seen as the “center” of this set). Let D_Π be the convex set of matrices with diagonal Π . The minimization set is then

$$C = S_N^+ \cap (I_N - S_N^+) \cap D_\Pi = \{0 \leq K \leq I_N, \text{diag}(K) = \Pi\}.$$

Set C is closed and convex as the intersection of closed convex sets. Because it is also bounded and non-empty and the space is finite dimensional, it is a compact set. As the objective function is continuous, the minimization problem admits (at least) one solution by Weierstrass Theorem. More precisely, the domain C is a projected spectrahedron, that is the projection of the spectrahedron H_+ , intersection of the cone of positive semidefinite matrices and an affine space. Indeed, the inequality $0 \leq K \leq I$ is equivalent to $\begin{pmatrix} K & 0 \\ 0 & I_N - K \end{pmatrix} \geq 0$. The minimization problem we consider here is then a particular case of the (non-linear) semidefinite optimization problems. The number of studies on semidefinite optimization has been growing rapidly (in the linear setting) since the 90's (see for instance Blekherman (2013), Vandenberghe (1996)).

The objective function can be rewritten using the indefinite norm $\|\cdot\|$ and the associated form $\langle\langle\cdot, \cdot\rangle\rangle$ defined previously. Recall that the indefinite norm is always nonnegative on the cone S_N^+ , and that $\langle\langle\cdot, \cdot\rangle\rangle$ is a scalar product for positive y .

PROPOSITION 4.1. *The minimization problem (P) is equivalent to the following problems:*

- (a) $\arg \min_{K \in C} \langle\langle I_N - K, K \rangle\rangle$;
- (b) $\arg \max_{K \in C} \|I_N - K\|^2 + \|K\|^2$, the “largest perimeter of the triangle $0KI$, $K \in C$ ”;
- (c) $\arg \max_{K \in C} \|K\|^2$, the “largest vector of C ”;
- (d) $\arg \max_{K \in C} \|K - \frac{1}{2}I_N\|^2$, “farthest vector (of C) to the center”;
- (e) $\arg \max_{K \in C} \|I_N - K\|^2$.

PROOF. By Proposition 3.2, $\text{MSE}(\hat{t}_{yHT}) = \langle\langle I - K, K \rangle\rangle$. We develop the inner product, and remark that $K \mapsto \langle\langle I, K \rangle\rangle$ is constant on C .

From now on, we restrict our study to the case of a positive variable y . In this case, $\langle\langle\cdot, \cdot\rangle\rangle$ is a scalar product, and we observe that the objective function is concave in K . The minimum is then attained at an extremal point of the convex set C . Moreover, as C is a compact convex set of \mathbb{R}^{N^2} it is the convex envelop of its extreme points by the Krein-Milman Theorem. Unfortunately, the extreme points of this set do not generally admit a simple characterization, even in the case $\sum_k \Pi_k$ is an integer (in particular, they may not be projections). The following simple particular case illustrates this point.

Let σ be the spectrahedron of positive semidefinite matrices of diagonal $\frac{1}{N}$. It is homothetic to the ellipsope ε of correlation matrices, that has been extensively studied (Ycart (1985), Grone (1990), Laurent (1995a), Laurent (1995b), Kurowicka (2003), Nagy (2014)). Let K be a matrix of σ . Its trace is 1, and as the eigenvalues are nonnegative, they are also less than 1, and K is a contracting matrix. Consider now the problem (P) with y any positive vector and C the set of contracting matrices with diagonal elements equal to $\frac{1}{N}$. The spectrahedron σ and the projected spectrahedron C coincide in this special case. Surprisingly, the optimisation problem (P) is very simple in this case, whereas the structure of σ is already complex. In fact, it is known that (for real matrices):

- THEOREM 12. (a) For any integer k such that $\binom{k+1}{2} \leq N$, there exists a matrix of rank k that is an extreme point of σ (Grone (1990) Theorem 2).
- (b) The edges of σ (extreme points where the normal cone to σ is of rank N) are the projections of σ (rank 1 matrices).
- (c) The linear optimization problem $\max_{K \in \sigma} \langle A, K \rangle$ is NP-hard, and the solution may not be an edge (projection).

These results show that the geometry of C is far from trivial, as is the linear minimization problem. It is also proved in the general case that extreme points of a projected spectrahedron satisfy rank constraint. Surprisingly, we have the following result:

PROPOSITION 4.2. The set of solutions of the problem (P) over C is $\{P \in C | P^T P = P\}$, the subset of projections.

The set of solution is then the set of edges of C , strict subset of its extreme points.

PROOF. We consider complex matrices, but the proof also works for real matrices. Set $K \in C$. As $|K_{k,l}|^2 \leq \frac{1}{N^2}$ then

$$\begin{aligned} \varphi(K) &= \sum_{k \in U} (w_k y_k)^2 - \sum_{k,l \in U} w_k y_k w_l y_l |K_{kl}|^2 \\ &\geq \sum_{k \in U} (w_k y_k)^2 - \frac{1}{N^2} \sum_{k,l \in U} w_k y_k w_l y_l, \end{aligned}$$

with equality iff all the off-diagonal elements are of modulus $\frac{1}{N}$.

Let P be a projection of C . As its trace equals 1, it is of rank 1 and there exists a vector $b \in \mathbb{C}^N$, $P = b\bar{b}^T$. Then $P_{k,k} = |b_k|^2 = \frac{1}{N}$ and $|P_{k,l}| = |b_k \bar{b}_l| = |b_k| \cdot |\bar{b}_l| = \frac{1}{N}$ for all $1 \leq k, l \leq N$, and P is a minimum of φ over C . Conversely, let K be a solution of (P). Then its off-diagonal elements are of modulus $\frac{1}{N}$, and the second order probabilities $\pi_{k,l}$ are equal to 0 for all $1 \leq k, l \leq N$. The determinantal sampling design with matrix K is of fixed size (equal to 1) and K is a projection by Corollary 1.1.

More generally, the same proof works for matrices whose diagonal elements sum to 1. The conclusion also holds for a constant variable y , whatever the vector Π whose sum of coefficients is an integer.

The previous results raise (open) questions for more complex spectrahedra (for vectors Π whose sum of coefficients is an integer). Are the minimizers always projections (resp. edges)? Or does the set of minimizers always contain a projection (resp. an edge)? Should this be the case, the conclusion would obviously also hold for Problem (P) with unconstrained diagonal (for designs of random size with integer expectation). Practically, we always found projection in our numerical studies.

5. Application

5.1. Principle

Finding the solution of a minimization problem such as (13) is cumbersome. The number of parameters is $N^2 + N$ and the space of minimization has a complex geometry. Therefore we carry out an application with a small population. For instance, this situation can be encountered in the case of the French household surveys. For those surveys, the sample is drawn according to a self-weighted two-stage sampling design. At the first stage, the primary units (PU) consist of urban areas or of the gathering of rural municipalities. The sampling design of the PU's is stratified with inclusion probabilities proportional to their number of dwellings. In the stratum of urban areas with 20000 to 100000 inhabitants of the Rhne-Alpes region, 6 UPs have to be drawn among 18. We set up various determinantal strategies for the selection of these 6 UPs, which we compare with one of their equivalent among the existing sampling designs. For each strategy we use the parametrization $K = \lambda_{max}^{-1} V^t V$, where V is a (N, p) real matrix, with $p \geq 6$, and λ_{max} is the largest eigenvalue of $V^t V$. By Corollary 1.1, this ensures that the random sample contains at least one element almost surely. We minimize criterion (14) as a function of V and w , for some auxiliary variables. The auxiliary variables are denoted by X . For each UP, they consist of the total numbers of dwellings, individual dwellings, social dwellings, inhabitants and owners of their principal residence.

- (a) **strategy 1:** K is a matrix whose values on the diagonal are equal to $\frac{1}{3}$. K is the best matrix obtained by the application of Theorem 5 to the $1^{st}, 5^{th}, 7^{th}$ 18-roots of unity, according to criterion (14) (Figure 1). The opponent within the existing designs is a SRS.
- (b) **strategy 2:** K is a matrix whose values on the diagonal are equal to $\frac{1}{3}$. K is obtained by minimizing criterion (14) for the set of variables X under the constraints $w_{kk} = \pi_k^{-1}$ and $(\lambda_{max}^{-1} V^t V)_{kk} = \frac{1}{3}$. The opponent is a balanced sampling design on variables X using the Cube Method (Deville and Tillé (2004)) with constant inclusion probabilities.
- (c) **strategy 3:** K is a matrix whose values on the diagonal are equal to $\frac{1}{3}$. K is obtained by minimizing criterion (14) for the set of variables X under the constraints $0.25 \leq w_{kk} \pi_k \leq 1.75$ and $(\lambda_{max}^{-1} V^t V)_{kk} = \frac{1}{3}$. The opponent is a balanced sampling design on variables X using the Cube Method (Deville and Tillé (2004)) with constant inclusion probabilities and the estimator is a calibration estimator (Deville and Sarndal (1992)) with the same set of variables.
- (d) **strategies 4,5,6** are the same as strategies 1, 2, 3 but with inclusion probabilities proportional to the number of dwellings.
- (e) **strategies 7,8** are the same as strategies 2, 3 or 5, 6. Nevertheless, the inclusion probabilities are no more fixed a priori. For the determinantal designs, they are given as a result of the optimization problem. These optimal probabilities are then used within the algorithm of Deville and Tillé (2004) to draw the comparison with the opponents.

Table 2. Application results

Values of $g(X)$			Median of $100t_{y^{qw}}^{-1}\sqrt{MSE(\hat{t}_{y^{qw}})}$		
	Method			Method	
Strategy	Determinantal	Opponent	Strategy	Determinantal	Opponent
1	82739.86	137105.74	1	17,46	16,73
2	24295.00	35024.40	2	7,53	7,48
3	3780.44	0	3	9,00	9,54
4	35646.31	36620.96	4	6,40	5,50
5	8601.15	13756.82	5	5,72	5,44
6	4461.48	0	6	6,26	11,40
7	5089.56	8225.72	7	6,22	9,09
8	1884.79	0	8	5,59	10,49

For the determinantal sampling designs, the value of $g(X)$ is exact using formula (1) of the MSE. For the opponents, except for the simple sampling design for which a variance formula is available, the values of $g(X)$ are obtained by Monte Carlo simulations. Since the population and sample sizes are small, only the linear distance provides, without failing, a calibration estimator for each simulation according to the method described in Deville and Sarndal (1992). This distance is then used for the opponents in strategies 3, 6 and 8.

5.2. Results

For each strategy, table 2 provides the optimum value of $g(X)$ using the determinantal sampling design and the value of $g(X)$ for the opponent. To evaluate the performance of each strategy for the estimation of the total of variables other than those used for the optimization, calibration or balanced sampling, table 2 provides the median of $100t_{y^{qw}}^{-1}\sqrt{MSE(\hat{t}_{y^{qw}})}$ for $Q = 70$ financial variables of interest.

The table gives rise to the following comments

- For the determinantal sampling design, the solution of the optimization problem depends highly on the initialization parameters. We then used a stochastic quasi-Newton method with several hundred of iterations to solve the problems.
- Although we have looked for the solutions among the matrices with a rank ranging from n to N , for each of the 8 strategies the best matrix found has always been a projection matrix.
- Moreover, for strategies 2, 7, 8 the optimal determinantal sampling design consists of a stratified sampling design with one unit to select among 6 strata of 3 units.
- Except for strategies 3, 6 and 8, the determinantal sampling design has better results than its opponent for the estimation of the total of auxiliary variables (table (2))
- As the result of overfitting for calibration with the Deville-Särndal method, the minimization with respect

to both the determinantal sampling and the weights leads to better results for the estimation of the total of intersets variables variables (table (2)).

- (f) A huge difference between determinantal sampling designs and their opponent is that second order inclusion probabilities are known exactly, which avoids the use of Monte Carlo methods for computing the variance.

6. Concluding remarks and perspectives

This study was conducted in the classical setting of sampling in a finite universe. However, it may be interesting in particular cases to sample in an infinite population, for instance in spatial statistics. Determinantal point processes are also adapted to this case, by replacing matrices with locally trace-class operators (see Soshnikov (2000), Hough (2009)).

Section 4 studies the classical problems of balanced sampling and calibration within the class of determinantal sampling designs, from the optimization point of view. This can be done using (10), that expresses the MSE of the estimator in terms of the matrix parameter K . This equation renders other approaches computable (at least numerically), for instance minimax procedures. Also, when no auxilliary variable are available, other cost functions can be chosen. For instance $\Theta(K, w) = -H(\mathcal{P}_K)$ searches sampling designs with maximal entropy (for instance within fixed-size determinantal sampling designs of fixed first order inclusion probabilities). It is a convex function. But its computation cost may be prohibiting.

Finally, we give a last look at Theorem 3. This Theorem claims that a determinantal sampling design is a mixture of fixed-size determinantal designs: one first performs Poisson sampling on the population of eigenvectors, and on a second stage, conditionnally on these eigenvectors, one performs a fixed-size determinantal sampling. Moreover, the Kernel K_I is just the Horvitz-Thompson estimator of K . From a statistical or algorithmic point of view, there is no need to restrict our attention to Poisson sampling at the first stage (but from a probabilistic point of view, we go outside the family of determinantal point processes). Also, there is no need to consider only the Horvitz-Thompson estimator. Such mixtures of determinantal point processes actually already appeared in the litterature in a particular case: these are the k -DPP of Kuleska and Taskar Kuleska (2011). These processes are defined as determinantal point processes conditionned to have exactly k elements. In Kuleska (2011), they prove that these processes are mixture of fixed-size determinantal point processes, where the eigenvectors are selected according to a k -Poisson sampling design, that is a Poisson sampling design conditionned to have exactly k elements. The k -Poisson sampling design is not determinantal.

To keep trace of the good properties of determinantal sampling, it could be intereting to study such mixtures where the selection of the sample of eigenvectors is determinantal. A possible construction is as follows. Let $K = \sum_{j \in \mathbb{U}} \lambda_j \phi_j \phi_j^*$ be a contracting matrix and J a contracting matrix with diagonal elements ν_i such that

$\lambda_i \leq \nu_i \leq 1$. We define the sampling design \mathcal{P}_{K_J} by: for all $s \in 2^U/\emptyset$

$$pr(s \subseteq \mathbb{S} | \mathbb{J} = J) = \det \left((K_J)_{|s} \right) \quad (15)$$

where $K_J = \sum_{j \in \mathbb{J}} \frac{\lambda_j}{\nu_j} \phi_j \phi_j^*$.

Even if this “2-determinantal” random sampling design is not determinantal in general, its inclusion probabilities are all computable. In addition, we can sample from K_J using Algorithm 1.1 twice. This provides a larger family of sampling designs that may deserve attention.

References

- Ben-Israel, A. and Greville, T. N. E. (2003) *Generalized inverses. Theory and applications (2nd ed.)*. New York, NY: Springer.
- Berger, Y. G. (1998) Rate of convergence for asymptotic variance of the Horvitz-Thompson estimator. *J. Statist. Plann. Inference* 74, no. 1, 149–168.
- Blekherman, G., Parrilo, P. A. and Thomas, R. R. (2013) *Semidefinite Optimization and Convex Algebraic Geometry*, SIAM.
- Borcea, J. B., Brandèn, P. and Liggett, T. M. (2009) Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.* 22, no. 2, 521–567.
- Borodin, A. (2011) Determinantal point processes. In *The Oxford handbook of random matrix theory*, pp. 231–249. Oxford Univ. Press, Oxford.
- Brandèn, P. and Jonasson, J. (2012) Negative dependence in sampling. *Scand. J. Stat.* 39 no. 4, 830–838.
- Cardot, H., Chaouch, M., Goga, C. and Labrune, C. (2010) Properties of design-based functional principal components analysis. *J. Statist. Plann. Inference* 140, no. 1, 75–91.
- Casazza, P. G., Fickus, M., Kovacevic, J., Leon, M. T. and Tremain, J. C. (2006) A Physical Interpretation for Finite Tight Frames. In *Harmonic Analysis and Applications, Applied and Numerical Harmonic Analysis* (C. Heil editor), pp. 51–76. Birkhuser Boston.
- Casazza, P. G., Redmond, D. and Tremain, J. C. (2008) Real Equiangular Frames, CISS Meeting, Princeton, NJ (2008).
- Cassel, C. M., Srndal, C. E. and Wretman, J. H. (1977) *Foundations of Inference in Survey Sampling*. New York: John Wiley & Sons.

- Chauvet, G., Bonnery, D. and Deville, J-C. (2011) Optimal inclusion probabilities for balanced sampling, J. Statist. Plann. Inference 141, no. 2, 984–994.
- Chauvet, G. (2014) A note on the consistency of the Narain-Horvitz-Thompson-Estimator arXiv:1412.2887v1
- Christensen, O. (2003) *An introduction to frames and Riesz bases*, Birkaser, Boston.
- Dhillon, I. S., Heath Jr, R. W., Sustik, M. A. and Tropp, J. A. (2005). Generalized finite algorithms for constructing Hermitian matrices with prescribed diagonal and spectrum, SIAM J. Matrix Anal. Appl. 27, 61–71.
- Deming, W. E. and Stephan F. F. (1941) On the interpretation of censuses as samples. J. Amer. Statist. Assoc. 36, 45–49.
- Deville, J-C. and Till, Y. (2004) Efficient balanced sampling: The cube method, Biometrika, 91, 893–912.
- Deville, J-C. and Särndal, C-E. (1992) Calibration estimators in survey sampling. J. Amer. Statist. Assoc. 87, no. 418, 376–382. Echantillonnageéquilibré exact poissonnien
- Deville, J-C. (2014) Echantillonnage équilibré exact poissonnien in Colloque Francophone sur les Sondages, Dijon, 2014.
- Dol, W., Steerneman, T. and Wansbeek, T. (1996) Matrix algebra and sampling theory: The case of the Horvitz-Thompson estimator, Linear Algebra Appl. 237/238, 225–238
- Erdös, P. and Rényi, A. (1959) On the central limit theorem for samples from a finite population. Magyar Tud. Akad. Mat. Kutat Int. Kzl. 4, 49–57.
- Grone, R. D., Pierce, S. and Watkins, W. (1990) Extremal correlation matrices. Linear Algebra Appl. 134, 63–70.
- Hajek, J. (1960) Limiting distributions in simple random sampling from a finite population. Magyar Tud. Akad. Mat. Kutat Int. Kzl. 5, 361–374.
- Hajek, J. (1964) Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Statist. 35 (1964), 1491–1523.
- Hajek, J. (1981) *Sampling from a finite population*. Marcek Dekker.
- Hedayat, A. S. and Sinha, K; K. (1991) *Design and inference in finite population sampling*. Wiley, New-York.
- Horn, A. (1954) Doubly stochastic matrices and the diagonal of a rotation matrix, American Journal of Mathematics 76, 620–630.
- Horn, A. and Johnson, C. R. (1991) *Topics in Matrix Analysis*. CUP.

- Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe, *J. Amer. Statist. Assoc.* 47, 663–685.
- Ben Hough, J. B., Krishnapur, M., Peres, Y. and Virag, B. (2009) *Zeros of Gaussian analytic functions and determinantal point processes*. University Lecture Series, 51. American Mathematical Society, Providence, RI.
- Ben Hough, J. B., Krishnapur, M., Peres, Y. and Virag, B. (2006) Determinantal point processes and independence. *Probability Surveys*, 3, 206–229.
- Isaki, C. T. and Fuller, W. A. (1982) Survey design under the regression superpopulation model. *J. Am. Statist. Ass.* 77, 89–96.
- Kadison, R. V. (2002) The Pythagorean theorem. I. The finite case. *Proc. Natl. Acad. Sci. USA* 99, no. 7, 4178–4184.
- de Klerk, E. (2002) *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*. Applied Optimization Series. Kluwer Academic, Boston, MA.
- Kulesza, A. and Taskar, B. (2011) k-DPPs: fixed-size determinantal point processes.
- Kurowicka, D. and Cooke, R. (2003) A parameterization of positive definite matrices in terms of partial correlation vines. *Linear Algebra Appl.* 372, 225–251.
- Lam, T. Y. (2001) *A first course in noncommutative rings*. Graduate Texts in Mathematics 131 (2nd ed.), Berlin, New York: Springer-Verlag.
- Laurent, M. and Poljak, S. (1995a) On a positive semidefinite relaxation of the cut polytope, *Linear Algebra Appl.* 223, 439–461.
- Laurent, M. and Poljak, S. (1995b) On the facial structure of the set of correlation matrices, *SIAM J. Matrix Anal. Appl.* 17, 530–547.
- Lavancier, F., Moller, J. and Rubak, E. (2015) Determinantal point process models and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 77, no. 4, 853–877.
- Lyons, R. (2003) Determinantal probability measures. *Publ. Math. Inst. Hautes Etudes Sci.* 98, 167–212.
- Macchi, O. (1975) The coincidence approach to stochastic point processes. *Adv. Appl. Probab.* 7, 83–122.
- Nagy, E., Laurent, M. and Varvitsiotis, A. (2014) Forbidden minor characterizations for low-rank optimal solutions to semidefinite programs over the elliptope. *J. Combin. Theory Ser. B* 108, 40–80.
- Narain, R. (1951) On sampling without replacement with varying probabilities. *J. Indian Soc. Agric. Statist.* 3 169–175.

- Neyman, J. (1934) On the Two Different Aspects of the Representative Methods: The Method of Stratified Sampling and the Method of Purposive Selection, *J. R. Stat. Soc.* 97, 558–625.
- Peirce, B. (1881) Linear associative algebra, *Amer. J. Math.* 4, 97–229.
- Pemantle, R. and Peres, Y. (2014) Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. *Combin. Probab. Comput.* 23, no. 1, 140–160.
- Robinson, P. M. (1982) On the Convergence of the Horvitz-Thompson Estimator. *Australian Journal of Statistics*, 24, 234–238.
- Sarndal, C-E., Swensson, B. and Wretman, J. (1992) *Model Assisted Survey Sampling*. Springer, New York.
- Scardicchio, A., Zachary, C. E and Torquato, S. (2009) Statistical properties of determinantal point processes in high dimensional Euclidian spaces. *Physical Review*, 79.
- Schur, J. (1911) Bemerkungen zur Theorie der beschrnkten Bilinearformen mit unendlich vielen Vernderlichen. *Journal fr die Reine und angewandte Mathematik (Crelle’s Journal)* 140 1–28.
- Soshnikov, A. (2000) Determinantal random point fields. *Russian Math. Surveys*, 55, 923–975.
- Soshnikov, A. (2002) Gaussian limit for determinantal random point fields. *Ann. Probab.* 30, no. 1, 171–187.
- Sen, A. R. (1953) On the estimate of the variance in sampling with varying probabilities. *J. Indian Soc. Agric. Statist.* 5, 119–127.
- Sustik, M. A., Tropp, J. A., Dhillon, I. S. and Heath Jr, R. W. (2007) On the existence of equiangular tight frames. *Linear Algebra Appl.* 426, no. 2–3, 619–635.
- Tropp, J. A. (2005) Complex equiangular tight frames. In: *Proceedings of SPIE 2005 (Wavelets XI)*, San Diego.
- Vandenberghe, L. and Boyd, S. (1996) Semidefinite Programming, *SIAM Review* 38, 49–95.
- Yates, F. and Grundy, P. M. (1953) Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc. Ser. B* 15, 235–261
- Ycart, B. (1985) Extreme points in convex sets of symmetric matrices. *Proc. Amer. Math. Soc.* 95, no. 4, 607–612.
- Yuan, M., Su, C. and Hu, T. (2003) A central limit theorem for random fields of negatively associated processes, *J. Theoret. Probab.* 16, 309–323.